An Unmanned Spacecraft Subsystem Cost Model for Advanced Mission Planning

G. Madrid, Jet Propulsion Laboratory, March 23, 1998

1. Introduction

As a NASA center, the Jet Propulsion Laboratory (JPL) is committed to the concept of developing and launching a continuously improving series of smaller robotic space exploration missions in shorter intervals of time (faster, better, cheaper). In order to plan and budget these advanced missions, JPL has begun an institutional initiative labeled “Develop New Products” (DNP). This institutional initiative involves an across the board paradigm shift in the process with which new projects are planned, designed, and implemented in an accelerated implementation cycle. A key factor in the planning of these missions is an accurate estimation of their cost so that an adequate, yet efficient, budget may be proposed that will not only be acceptable to NASA but will ensure a realistic implementation of a specific project within a predetermined project implementation schedule and risk envelope.

The project planning process has also been accelerated so that cost estimates may be produced within a one to two week cycle. This permits a second or third cost estimate to be produced that takes into account technology-cost trades vs. science objectives derived from the advanced planning deliberations in which the cost estimators play a key role. Once converged, this process leads to a budget estimate that has achieved a certain degree of consensus within the JPL community and its industrial partners prior to entering the proposal stage. Because of this, the probability of approval of the proposal is greatly increased.

The main instrument for carrying out this advanced planning process is a team of spacecraft and ground system engineering experts termed “Team X” at JPL. The team members are key technical staff selected by the JPL technical divisions as having the expertise required to design and cost the subsystem to which they have been assigned. This team conducts its deliberations around a distributed workstation facility that interacts through a network in conjunction with a central database and a documentarian. This arrangement permits the study leader and team members to interact in “real time” to develop a preliminary design and cost estimate within a week. Such a process would normally have taken from three to four months under the previous paradigm. A large portion of the proposals reviewed by Team X are of the DNP type. In order for a new project to be termed “DNP”, the proposal must establish that the implementation (from concept to launch) can be accomplished within 33 months and the final cost estimate must fall into a cost range between $120M and $500M, not including the launch vehicle. Projects falling outside this range are processed using other more pertinent models.
are examined by the Study Lead and the Team X system engineer and may be overridden by them.

The cost estimation process uses differing approaches to predicting cost based on the portion of the work breakdown structure (WBS) being estimated. The basic methods used for estimating the cost of the distinct portions of the total project cost are:

1. Statistically-based algorithms from the previous Deep Space Cost Model that have been adjusted to conform to the DNP paradigm. This type of algorithm are termed historically-based algorithms (Hist. Based Algo.)
2. A non-statistical algorithm based on a quasi-grass-roots-based estimate and expert opinion formulated in consultation with technical specialists in the area of the project component being assessed. The algorithm is based on an evaluation of actual data and the design of the function being performed but which does not have sufficient structure to formulate a model at this time.
3. The current Instrument Cost Model developed by Keith Warfield
4. The current Subsystem Cost Model developed by Leigh Rosenberg
5. The actual price of the item being assessed, as in the case for launch vehicles, where the cost to the government is either predetermined by agreement with the vendors or is the listed price for the service.

The following lists the components of the advanced project cost estimation process and the method used:

<table>
<thead>
<tr>
<th>Project Cost Component</th>
<th>Cost Est. Method</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Project Management and Administration</td>
<td>Hist.-Based Algo.</td>
</tr>
<tr>
<td>2. Science and Science Team Activities</td>
<td>Quasi-GR-Based Algo.</td>
</tr>
<tr>
<td>3. Project and Mission Engineering</td>
<td>Hist.-Based Algo.</td>
</tr>
<tr>
<td>5. Spacecraft (System &amp; Subsystem Costs)</td>
<td>Hist.-Based Algo.</td>
</tr>
<tr>
<td>5.1 System Level Mgmt &amp; Engrg</td>
<td>S/S Cost Model</td>
</tr>
<tr>
<td>5.2 S/C Subsystem Costs</td>
<td>Quasi-GR-Based Algo.</td>
</tr>
<tr>
<td>6. Assembly, Test, and Launch Operations</td>
<td>Quasi-GR-Based Algo.</td>
</tr>
<tr>
<td>8. Launch Vehicle</td>
<td></td>
</tr>
</tbody>
</table>

The discussion in this paper concerns itself solely with the spacecraft subsystem costs, item 5.2. Mr. Rosenberg’s paper will discuss the overall process (items 1-8) while Mr. Warfield’s paper will deal with the instrument model used in item 4.
This paper describes the subsystem portion of the unmanned mission and spacecraft implementation cost model used in this interactive environment that is consistent with the DNP assumptions. This mission and spacecraft subsystem cost model was developed by Mr. Leigh Rosenberg of JPL. An adjunct instrument model was developed by Mr. Keith Warfield, also of JPL. Companion papers are being submitted by Mr. Rosenberg, Mr. Warfield, and other cost team members that describe other aspects of the new cost estimation environment including the historical and evolutionary aspects. The focus of this paper will be on the design and structure of the subsystem cost model itself.

2. Model Overview

Because no unmanned space missions have yet been fully implemented using the new spacecraft development lifecycle paradigm shift, the cost model used is not based on a historical data base of previously implemented missions. Rather, the model is based on a data base of the prior estimates of proposed missions that have been developed using the Team X process and that have been certified as viable candidates for future mission proposals. As a result, the model described here acts as a predictor of Team X results and is currently used to validate the on-going estimates being developed with respect to a consistency with the DNP Process, past predictions, and previously proposed designs.

The focus of this paper is on a subsystem cost model that is based on data obtained from the Team X process, not on the process or estimates obtained by the team. Although the model is a predictor of the planning team results, it was nonetheless designed as if the parameters and cost data were obtained from an as-built design. An effort is under way to validate model estimates obtained using the new paradigm as soon as mission implementation costs are available from more recent missions that do business under the new paradigm.

The Cost Model is linked to the Team X system and subsystem workstations so that the technical parameters required by the model are passed to the cost workstation which updates the estimates of the cost for each subsystem as the deliberations are in progress. The model cost estimates are then used as a comparator to the costs being estimated by the team and are kept separate from the team deliberations so as not to bias the results. The Model cost estimates used in this manner are calculated using algorithms derived from the statistical analysis performed on the data base of DNP projects mentioned above.

Some of the non-technical project/system infrastructure costs used during the Team X sessions are estimated by algorithms derived from historical costs for similar type projects (scaled to the DNP project time phase constraints). Since they are a function of total system, subsystem, and instrument costs, the algorithms permit a quick assessment of the infrastructure costs as the subsystem costs are being developed. At the end of the deliberations the predicted infrastructure costs

3. *Cost Model Data Base*

The Subsystem Cost Data Base is a collection of all of the system and subsystem technical parameters, subsystem masses, and associated cost estimates obtained as the result of Team X deliberations from October 1996 through October 1997. Of the nineteen proposed unmanned deep space projects whose estimates and parameters are contained in the data base, seventeen have been selected for application for the cost model. Other project cost estimates produced by Team X during the period the data base was constructed were excluded due to their unique characteristics which did not entirely fit into the DNP mold. The data base parameters are continuously undergoing some fine tuning as Team X review of the design, results in modification to the parameters.

Table 1, below, lists the cost portion of the data base by project. Due to the sensitive nature of the cost data regarding projects, these are only identified by a placeholder identification as P1, P2, etc.

**Table 1. Subsystem Data Base Cost Summary**

<table>
<thead>
<tr>
<th>Project</th>
<th>Tot$M</th>
<th>ADCS</th>
<th>C&amp;DH</th>
<th>Telcom</th>
<th>Power</th>
<th>Prop</th>
<th>Struct</th>
<th>Therm</th>
<th>Core</th>
<th>MeB/U</th>
<th>Other</th>
</tr>
</thead>
<tbody>
<tr>
<td>P1</td>
<td>91.4</td>
<td>17.8</td>
<td>12.7</td>
<td>2.0</td>
<td>15.0</td>
<td>6.7</td>
<td>15.7</td>
<td>12.3</td>
<td>3.4</td>
<td></td>
<td>5.8</td>
</tr>
<tr>
<td>P2</td>
<td>96.7</td>
<td>17.7</td>
<td>9.1</td>
<td>1.4</td>
<td>13.0</td>
<td>14.6</td>
<td>19.8</td>
<td>9.5</td>
<td>3.4</td>
<td></td>
<td>8.4</td>
</tr>
<tr>
<td>P3</td>
<td>95.0</td>
<td>13.4</td>
<td>4.4</td>
<td>2.0</td>
<td>13.9</td>
<td>15.2</td>
<td>20.7</td>
<td>10.3</td>
<td>3.4</td>
<td></td>
<td>2.2</td>
</tr>
<tr>
<td>P4</td>
<td>67.9</td>
<td>4.0</td>
<td>3.5</td>
<td>1.7</td>
<td>8.2</td>
<td>3.4</td>
<td>4.5</td>
<td>1.7</td>
<td>1.7</td>
<td></td>
<td>9.5</td>
</tr>
<tr>
<td>P13</td>
<td>69.2</td>
<td>17.8</td>
<td>8.5</td>
<td>1.0</td>
<td>10.2</td>
<td>4.6</td>
<td>4.1</td>
<td>8.9</td>
<td>2.8</td>
<td></td>
<td>1.7</td>
</tr>
<tr>
<td>P14</td>
<td>54.8</td>
<td>11.9</td>
<td>2.4</td>
<td>0.8</td>
<td>10.4</td>
<td>5.5</td>
<td>10.2</td>
<td>8.3</td>
<td>3.2</td>
<td></td>
<td>2.1</td>
</tr>
<tr>
<td>P15</td>
<td>33.4</td>
<td>6.1</td>
<td>2.1</td>
<td>0.6</td>
<td>5.0</td>
<td>5.3</td>
<td>3.5</td>
<td>7.4</td>
<td>1.7</td>
<td></td>
<td>1.7</td>
</tr>
<tr>
<td>P16</td>
<td>51.7</td>
<td>10.2</td>
<td>2.9</td>
<td>0.8</td>
<td>8.1</td>
<td>6.1</td>
<td>9.7</td>
<td>9.4</td>
<td>2.8</td>
<td></td>
<td>1.7</td>
</tr>
<tr>
<td>P17</td>
<td>36.7</td>
<td>6.2</td>
<td>2.4</td>
<td>0.7</td>
<td>5.4</td>
<td>5.8</td>
<td>3.5</td>
<td>8.2</td>
<td>2.8</td>
<td></td>
<td>1.7</td>
</tr>
<tr>
<td>Avg</td>
<td>71.1</td>
<td>12.3</td>
<td>5.9</td>
<td>1.2</td>
<td>10.1</td>
<td>12.5</td>
<td>10.8</td>
<td>9.6</td>
<td>3.1</td>
<td></td>
<td>3.1</td>
</tr>
<tr>
<td>Std Dev</td>
<td>20.6</td>
<td>4.6</td>
<td>2.8</td>
<td>0.4</td>
<td>3.2</td>
<td>9.2</td>
<td>5.6</td>
<td>2.1</td>
<td>0.7</td>
<td></td>
<td>1.8</td>
</tr>
<tr>
<td>Max</td>
<td>98.2</td>
<td>22.3</td>
<td>12.7</td>
<td>2.0</td>
<td>15.0</td>
<td>42.4</td>
<td>20.7</td>
<td>13.5</td>
<td>5.0</td>
<td></td>
<td>8.4</td>
</tr>
<tr>
<td>Min</td>
<td>33.4</td>
<td>6.1</td>
<td>2.1</td>
<td>0.6</td>
<td>5.0</td>
<td>4.6</td>
<td>3.4</td>
<td>4.5</td>
<td>1.7</td>
<td></td>
<td>1.7</td>
</tr>
</tbody>
</table>

Table 2, lists the instances of the design parameters, \( \xi \), which have been selected as having a causal relationship to cost for all projects in the data base.
The subsystem mass plays a role as a cost estimation parameter in some instances. Table 3 lists the subsystem mass data in the data base. When applicable to a particular regression fit, the subsystem mass is used as one of the technical parameters for the regression fit.

### Table 3. Subsystem Data Base Values for Mass

<table>
<thead>
<tr>
<th>Project</th>
<th>Mass Values by Subsystem for each Project (kg)</th>
<th>Total Cost/Cost (kg)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>ADCS</td>
<td>CADH</td>
</tr>
<tr>
<td>P1</td>
<td>25.7</td>
<td>14.6</td>
</tr>
<tr>
<td>P2</td>
<td>37.5</td>
<td>17.5</td>
</tr>
<tr>
<td>P3</td>
<td>18.7</td>
<td>8</td>
</tr>
</tbody>
</table>

| Avg     | 21.7  | 14.6  | 30.3    | 27.4  | 118.7  | 173.6 | 47      | 17.6       | 136.5      | 1044.2     | 5.5      | 699       |
| Std Dev | 5.7   | 3.7   | 15.5    | 69.6  | 141.7  | 112.7 | 18.5    | 31         | 71.1       | 420.5      | 0        | 699       |
| Max     | 350.0 | 619.3 | 324.0   | 88.9  | 110.0  | 600.0 | 30.0    | 300.0      | 800.0      | 4000.0     | 5.0      | 1029      |
| Min     | 0.0   | 0.0   | 0.0     | 0.0   | 0.0    | 0.0   | 0.0     | 0.0        | 0.0        | 0.0        | 0.0      | 0.0       |

4. **Model Construction**

In order to predict subsystem costs from the data presented in the data base, a model that relates subsystem cost to the parameters, \( \xi \), in Table 2 is required. The approach taken was to define a regression model function that could be used for each of the subsystems to predict cost within the parameter data domain. The cost data and the parameters relevant to each subsystem would form the basis for a first order regression fit that would result in an equation that would then be:

\[ \text{Cost} = a_0 + a_1 \text{Parameter}_1 + a_2 \text{Parameter}_2 + \ldots + a_n \text{Parameter}_n \]

G. Madrid

-5-

March 23, 1998
used to predict costs for that subsystem within the predictive constraints imposed by the fit. The total subsystem costs would then be obtained by summing all of the subsystem cost estimates.

A generalized first order multivariate linear regression function [Draper and Smith,1966, § 5.1] was used throughout. Although some of the relationships are non-linear, they may be transformed to this linear form (i.e., they are intrinsically linear). It was determined, through analysis and experimentation, that this approach would provide very acceptable fits for the data set currently in the database. This type of function is traditionally expressed as follows:

\[ \eta_i = \beta_0 + \sum \beta_j X_{ij} \quad (j=1,k) \]  

where \( \eta_i \) is the dependent variable, \( X_{ij} \) are the independent variables, \( \beta_j \) are the undetermined coefficients of \( X_{ij} \) to be determined by means of the linear regression process, and \( \beta_0 \) is a constant (also to be determined). The index, \( i \), refers to a particular instance where a measurement of \( \eta_i \) occurs for the specific subsystem for which the linear estimation is being made.

Assume that \( Y_i \) is the measurement of \( \eta_i \) such that,

\[ Y_i - \eta_i = \varepsilon_i \]  

where \( \varepsilon_i \) is the measurement error and errors are assumed to be additive and satisfy the Gauss-Markov assumptions [Beck and Arnold,1977, § 5.1.3].

This being the case, we may then express the regression function (4.1) as:

\[ Y_i = \beta_0 + \sum \beta_j X_{ij} + \varepsilon_i \quad (j=1,k) \]  

In the particular application in question, the following interpretation will be given to the variables and coefficients:

- \( Y_i \): The instance, \( i \), of a cost measurement, \( Y \), for the subsystem under assessment. \( Y_i \) is considered an estimate of the regression function, \( \eta_i \), of the parameter values \( (X_{ij}) \) pertaining to the specific instance.
- \( X_{ij} \): Instances of the technical parameters selected from the set \( \{\xi\} \) that
have a causal relation to the cost, Y. The selected parameters are ordered from \( j=1, k \), in the equation (4.3). This ordinal specification may be different than that used in the global set of parameters \( \{ \xi_i \} \) since only the parameters influencing the cost are selected.

The coefficients of the linear regression equation for the subsystem being assessed that are to be estimated by means of the regression process.

\[ \beta_j \]

The measurement error, \( \eta_i \).

This form (4.3), is the regression model to be used in the discussion that follows. Other model approaches (including non-linear) were examined but did not produce significant improvements in fit for the particular set of data being evaluated.

The linear regression estimation process operates on two sets of data defined from the data base. These are: 1) an nx1 matrix of the cost instances, \( Y_i \), for the subsystem being assessed, and 2) a corresponding nx(k+1) matrix of the instances of the technical parameters, \( X_{ij} \) selected as being causal for this subsystem.

\[
P_{4.3} \quad \beta_{j}
\]

\[ \epsilon_i \]

This form (4.3), is the regression model to be used in the discussion that follows. Other model approaches (including non-linear) were examined but did not produce significant improvements in fit for the particular set of data being evaluated.

The linear regression estimation process operates on two sets of data defined from the data base. These are: 1) an nx1 matrix of the cost instances, \( Y_i \), for the subsystem being assessed, and 2) a corresponding nx(k+1) matrix of the instances of the technical parameters, \( X_{ij} \) selected as being causal for this subsystem.

\[
Y = \begin{bmatrix}
  Y_1 \\
  Y_2 \\
  \vdots \\
  Y_n \\
\end{bmatrix}
\]

(4.4)

\[
X = \begin{bmatrix}
  1 & X_{11} & X_{12} & \cdots & X_{1k} \\
  1 & X_{21} & X_{21} & \cdots & \vdots \\
  1 & X_{31} & X_{32} & \cdots & X_{33} \\
  \vdots & \vdots & \vdots & \ddots & \vdots \\
  1 & \cdots & \cdots & \cdots & X_{nk} \\
\end{bmatrix}
\]

(4.5)

Using these data as input, the linear estimation process solves for estimates of \( \beta_j \) that minimize \( \epsilon_i \). These estimated coefficients are termed \( b_j \). In general, the results of the regression estimation is expressed with the predictive equation:

\[
y_i = b_0 + \sum_{j=1}^{k} b_j x_{ij} \\
(j=1,k)
\]

(4.6)
where $y_i$ is the predicted cost for the subject subsystem at any instance $i$, based on the estimated parameter coefficients, $b_j$, and the parameters, $x_{ij}$, specified for that subsystem at that instance. When using the predictive equation, care must be taken to ensure that the parameters selected fall within the domain of the data base parameters.

When the subsystem costs have been individually estimated, the total spacecraft system costs may be calculated by summing the subsystem results. Additional costs for system management, system engineering, spares, integration and test, and operations support need to be added to complete the cost estimate for the spacecraft. These costs and the costs associated with the project infrastructure itself will be dealt with in a follow-on paper.

The basic process in construction of the model were as follows:

1) Validate the model data base to ensure that all of the information is appropriate and accurate.
2) In consultation with subsystem technologists, establish the initial set of parameters, $X_{ij}$, casually related to estimating the cost of each subsystem $Y_i$ (e.g., mass, power generation, radiation dosage, etc.). Ensure that these are appropriately and accurately represented in the data base.
3) Determine the general regression function to be used (as above),
4) Conduct an evaluation strategy using the regression strategy selected to determine the "best" parameters to leave in the fit. In this case a modified backward elimination process was performed to reduce the set of parameters, $X_{ij}$ to be considered to those resulting in a validated "best fit" and whose t statistics indicate validate the hypothesis that $E(bj)=0$, consistent with a maximization of the Coefficient of Multiple Determination, $(R^2)$. Standard F- and t-test constraints for fit and coefficient validity were utilized.
5) Validate the resulting model against expected behavior within the valid range of the parameters. The model behavior is checked against independent subsystem estimates provided by the expert for that subsystem.
6) Reconstruct any of the model equations based on any new information obtained in the process of validating the model equation in (5).
7) The entire set of subsystem costs are then validated against the data base itself to ensure that the difference of the costs obtained vs. the data base costs for a particular project are within the expected variance of the model.

The current model equations will be updated as improved interpretation of the

G. Madrid

March 23, 1998
technical parameters is obtained by working with the technical experts in that area. The model equations will also be reviewed and validated as soon as actual cost data is available for DNP-Type projects. Work is in progress to collect cost and technical data from new projects as they enter the implementation stage so that the model may be validated or corrected with improved or actual cost information.

5. Linear Estimation Process and Resulting Statistics
The Ordinary Least Squares (OLS) method was selected to estimate the parameters. OLS is usually recommended when nothing is known about the measurement errors [Beck and Arnold, 1977, § 6.2], since even with little or no information on the error distribution, an adequate predictor may be obtained. However, when information regarding the statistics of the errors is known or assumed, the process produces an efficient estimator of the coefficients ($\beta_j$). This section analyzes the statistical results of the use of this method and identifies the general form of the predictive equation which is the basis for the Cost Estimation Relationships (CER's) which are discussed in the next section.

In order to be succinct in expressing the logic of the process, we will resort to matrix notation in describing the analysis [Beck and Arnold, 1977, § 6.2]. The sum of squares function used for ordinary least squares with the linear model $\eta = \mathbf{X}\beta$ is

$$ S = (\mathbf{Y} - \mathbf{X}\beta)^T (\mathbf{Y} - \mathbf{X}\beta) $$

(5.1)

where $\mathbf{Y}$ and $\mathbf{X}$ are defined by (4.4) and (4.5) respectively and $\beta$ is a vector of the undetermined coefficients $\beta_j$, where, $j=0,n$.

Assume that $b$ is the estimate of $\beta$. Then, since $\mathbf{Y}$ is the estimate of $\eta$ that is sought,

$$ \mathbf{Y} = \mathbf{X} \mathbf{b} $$

(5.2)

In order to solve for the estimated coefficients, $\mathbf{b}$, it is necessary to pre-multiply by $\mathbf{X}^T$ so that

$$ \mathbf{X}^T \mathbf{Y} = \mathbf{X}^T \mathbf{X} \mathbf{b} $$

(5.3)

Further pre-multiplication by $(\mathbf{X}^T \mathbf{X})^{-1}$, yields

G. Madrid  

March 23, 1998
\[
(X^T X)^{-1}X^T Y = (X^T X)^{-1}(X^T X) b = b
\] (5.4)

This results in the estimator of the coefficients \( \beta \)

\[
b = (X^T X)^{-1} X^T Y
\] (5.5)

It can readily be demonstrated that 5.5 minimizes the sum of squares, 5.1, and is the OLS estimator of the coefficients \( \beta_j \) [Johnson and Wichern, 1988, § 7.3].

For unique estimation of all the coefficients, \( \beta_j \), the matrix \( (X^T X) \) must be non-singular. This means that any one column in \( X \) cannot be proportional to any other column or any linear combination of columns because if such a proportionality exists the determinant of \( (X^T X) \) must equal zero.

As we have mentioned before, if the errors are additive, of zero mean in \( Y \) and \( X \), and the \( \beta \) are nonstochastic, then \( \text{E}(b) \) is an unbiased estimator of \( \beta \) such that,

\[
\text{E}(b) = (X^T X)^{-1} X^T X \beta = \beta
\] (5.6)

The covariance matrix for the coefficients is expressed as:

\[
\text{cov}(b) = (X^T X)^{-1} X^T \psi X (X^T X)^{-1}
\] (5.7)

where, \( \psi = (\varepsilon \varepsilon^T) = \sigma^2 I \).

If it is further assumed that the errors are uncorrelated and of constant variance, then the covariance matrix for the coefficients may be reduced to the expression,

\[
\text{cov}(b) = (X^T X)^{-1} \sigma^2
\] (5.8)

which is the minimum covariance matrix of \( b \). The variances of the \( \beta_j \) (or the \( \text{SE}^2 \), depending on the assumptions being used) may be obtained from the diagonal elements of this matrix [Draper and Smith, 1966, § 4.2].

Similarly, from the relationships, 5.5, 5.6, and 5.7, all of the necessary items required to evaluate the fit are obtained. The following table lists the basic data items needed for the analysis:
Table 4. Results of Linear Estimation Process Required for Assessment

<table>
<thead>
<tr>
<th>Estimated Coefficients (bj) and related statistics</th>
</tr>
</thead>
<tbody>
<tr>
<td>b_k</td>
</tr>
<tr>
<td>Est. Value (bj)</td>
</tr>
<tr>
<td>Std. Error (SE)</td>
</tr>
<tr>
<td>t Statistic</td>
</tr>
</tbody>
</table>

The predicted bj values and the standard errors for the coefficients are, of course, produced as a direct result of the least squares minimization. For the assumptions on the error being used, the following statements hold,

\[ E(b_j) = \beta_j \quad (5.10) \]

\[ V(b_j) = c_{jj} \sigma^2 \quad (5.11) \]

where the c_{jj} are the diagonal elements of \((X^TX)^{-1}\).

If \( \sigma \) is not known or normality is suspect then

\[ \text{est. var} (b_j) = c_{jj} s^2 \quad (5.12) \]

and,

\[ \text{SE}(b_j) = (c_{jj} s^2)^{1/2} \quad (5.13) \]

where \( s^2 \) is the sample variance for each bj and \( \text{SE}(b_j) \) is the standard error of estimate.

In a similar manner the variance of \( Y \), \( V(Y) \), or the standard error of \( Y \), \( \text{SE}(Y) \), can be determined from the diagonal elements of,

\[ \text{cov}(Y') = X (X^TX)^{-1}X^T \sigma^2 \quad (5.14) \]

Under the assumptions being invoked, the t statistic for each bj may be computed as,
\[ t_j = \frac{E(b_j)}{SE(b_j)} \]  

(5.15)

The Coefficient of Multiple Determination, \( R^2 \), is defined as,

\[ R^2 = SS_{\text{reg}} + SS_{\text{tot}} = \Sigma (Y'_i - \bar{Y})^2 + \Sigma (Y_i - \bar{Y})^2 \]  

(5.16)

where, \( SS_{\text{reg}} \) is the regression sum of squares (the deviation between the regression line \( Y'_i \) and the mean \( \bar{Y} \)) and \( SS_{\text{tot}} \) is the total sum of squares (the total deviation between the data \( Y_i \) and the mean \( \bar{Y} \)). However, since \( SS_{\text{tot}} \) is the sum of \( SS_{\text{reg}} \) and \( SS_{\text{resid}} \), the \( R^2 \) statistic may be calculated as,

\[ R^2 = SS_{\text{reg}} + (SS_{\text{reg}} + SS_{\text{resid}}) \]  

(5.17)

where, \( SS_{\text{resid}} \) is defined as \( \Sigma (Y'_i - Y_i)^2 \)

The F statistic, used in the test for lack of fit is computed as,

\[ F(df) = \frac{[SS_{\text{reg}} + k]}{SE(Y)^2} \]  

(5.18)

The F statistic for the fit is dependent on the degrees of freedom, \( df \), which is defined for the table above, as: the number of data points, \( n \), less the number of variables being determined in the regression analysis, \( k \) (including the constant, \( b_0 \)).

The F-test criteria for goodness of fit used is that,

\[ F > F_{\text{crit}} \]  

(5.19)

where \( F_{\text{crit}} \) is the \( F(k, df, \alpha) \) critical value from the F-tables. The greater \( F \) is than the \( F_{\text{crit}} \) value, the better the confidence that the "best" fit has been achieved.

6. Cost Estimation Relationships and Constraints

The cost estimation relationships, which are the direct expression of the model, are built utilizing the predictive equation (4.6), the coefficients determined in the linear estimation process, and the corresponding statistics described in section 5. This section summarizes the CER's developed for the Spacecraft Subsystem Model by subsystem, including the constraints imposed by
6.1. **Attitude Determination and Control (ADCS)**

The following CER for the estimated subsystem cost \( (Y) \) in millions of dollars (FY97) was determined for ADCS subsystems within the range of the data domain:

\[
Y = b_0 + b_1 X_1 + b_2 X_2 + b_3 X_3
\]  

(6.1)

### Coefficients & Constraints for ADCS

<table>
<thead>
<tr>
<th>X Parameters</th>
<th>units</th>
<th>Avg</th>
<th>S.Dev</th>
<th>Max</th>
<th>Min</th>
</tr>
</thead>
<tbody>
<tr>
<td>X0 Constant</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
</tr>
<tr>
<td>X1 Subsystem Mass</td>
<td>kg</td>
<td>16.06</td>
<td>13.6</td>
<td>48.1</td>
<td>1.9</td>
</tr>
<tr>
<td>X2 D/L Data Rate</td>
<td>kbps</td>
<td>60.63</td>
<td>149</td>
<td>600</td>
<td>0</td>
</tr>
<tr>
<td>X3 Pointing Knowledge</td>
<td>arcsecs</td>
<td>327</td>
<td>302</td>
<td>900</td>
<td>5</td>
</tr>
</tbody>
</table>

### Coeff. Values

<table>
<thead>
<tr>
<th>Coeff. Symbol</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>( b_0 )</td>
<td>9.674</td>
</tr>
<tr>
<td>( b_1 )</td>
<td>0.2428</td>
</tr>
<tr>
<td>( b_2 )</td>
<td>0.0064</td>
</tr>
<tr>
<td>( b_3 )</td>
<td>-0.004</td>
</tr>
</tbody>
</table>

6.2. **Command and Data Handling (C&DH)**

The following CER for the estimated subsystem cost \( (Y) \) in millions of dollars (FY97) was determined for C&DH subsystems within the range of the data domain:

\[
Y = b_0 + b_1 X_1 + b_2 X_2
\]  

(6.2)

### Coefficients & Constraints for C&DH

<table>
<thead>
<tr>
<th>X Parameters</th>
<th>units</th>
<th>Avg</th>
<th>S.Dev</th>
<th>Max</th>
<th>Min</th>
</tr>
</thead>
<tbody>
<tr>
<td>X0 Constant</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
</tr>
<tr>
<td>X1 D/L Data Rate</td>
<td>kbps</td>
<td>71.8</td>
<td>1.55</td>
<td>600</td>
<td>0</td>
</tr>
<tr>
<td>X2 Redundancy</td>
<td>ordinal</td>
<td>2.6</td>
<td>0.7</td>
<td>3</td>
<td>1</td>
</tr>
</tbody>
</table>

### Coeff. Values

<table>
<thead>
<tr>
<th>Coeff. Symbol</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>( b_0 )</td>
<td>0.3078</td>
</tr>
<tr>
<td>( b_1 )</td>
<td>0.0163</td>
</tr>
<tr>
<td>( b_2 )</td>
<td>2.4886</td>
</tr>
</tbody>
</table>

This CER covers the sum of both hardware and software for the C&DH subsystem.

6.3. **Telecommunications (Telecom)**

The following CER for the estimated subsystem cost \( (Y) \) in millions of dollars (FY97) was determined for Telecommunications subsystems within the range of the data domain:

\[
Y = b_0 + b_1 X_1 + b_2 X_2 + b_3 X_3 + b_4 X_4
\]  

(6.3)

G. Madrid
March 23, 1998
6.4. **Power Generation (Power)**

The following CER for the estimated subsystem cost \(Y\) in millions of dollars (FY97) was determined for Power subsystems within the range of the data domain:

\[
Y = b_0 + b_1 X_1 + b_2 X_2 + b_3 X_3 + b_4 X_4 + b_5 X_5 \quad (6.4)
\]

---

6.5. **Propulsion**

The following CER for the estimated subsystem cost \(Y\) in millions of dollars (FY97) was determined for Propulsion subsystems within the range of the data domain:

\[
Y = b_0 + b_1 X_1 + b_2 X_2 \quad (6.5)
\]
6.6. Structures

The following CER for the estimated subsystem cost (Y) in millions of dollars (FY97) was determined for Structures subsystems within the range of the data domain:

\[ Y = b_0 + b_1 X_1 + b_2 X_2 \]  

where \( Y = \ln (\text{cost}) \).

Coefficients & Constraints for Structures

<table>
<thead>
<tr>
<th>X Parameters</th>
<th>units</th>
<th>Avg</th>
<th>S.Dev</th>
<th>Max</th>
<th>Min</th>
</tr>
</thead>
<tbody>
<tr>
<td>X0 Constant</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
</tr>
<tr>
<td>X1 Ln (S/S Mass)</td>
<td>kg</td>
<td>72.9</td>
<td>59</td>
<td>220.1</td>
<td>7.4</td>
</tr>
<tr>
<td>X2 Ln ISP</td>
<td>n/a</td>
<td>6.1</td>
<td>1</td>
<td>8.2</td>
<td>5.4</td>
</tr>
</tbody>
</table>

Cost is obtained from this CER by computing \( e^Y \).

A supplementary estimate of the mechanical build-up that is usually associated with structures. This CER is,

\[ Y = b_0 + b_1 X_1 + b_2 X_2 \]  

6.7. Thermal Protection

The following CER for the estimated subsystem cost (Y) in millions of dollars (FY97) was determined for Power subsystems within the range of the data domain:

G. Madrid

March 23, 1998
\[ Y = b_0 + b_1 X_1 + b_2 X_2 \]  
(6.7)

### Coefficients & Constraints for Thermal

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Units</th>
<th>Avg</th>
<th>S.Dev</th>
<th>Max</th>
<th>Min</th>
</tr>
</thead>
<tbody>
<tr>
<td>X0 Constant</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
</tr>
<tr>
<td>X1 Redundancy</td>
<td>ordinal</td>
<td>0.8</td>
<td>0.4</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>X2 Active/Passive</td>
<td>ordinal</td>
<td>0.1</td>
<td>0.3</td>
<td>1</td>
<td>0</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>(b_0)</td>
<td>1.817</td>
</tr>
<tr>
<td>(b_1)</td>
<td>1.068</td>
</tr>
<tr>
<td>(b_2)</td>
<td>4.255</td>
</tr>
</tbody>
</table>

### 6.8. Statistical Summary

In evaluating each CER the statistics on the \(b_j\) coefficients and the estimated response variable, \(Y\) were analyzed. The \(t\) statistics were tested to determine if the resulting estimates for the coefficients were significant contributors. This information was used in determining which coefficients to leave in the regression estimate and which to drop out. In general, the final \(t\) statistics satisfied the \(t\)-test criteria for significance. The \(R^2\) and the \(F\) statistic were used to determine the goodness of fit of the resulting predictive equation for \(Y\). The following table summarizes the estimate statistics associated with the CER's listed above.

**Table 5. Summary Estimate Statistics**

<table>
<thead>
<tr>
<th>Subsystem</th>
<th>(R^2)</th>
<th>(F)</th>
<th>(k)</th>
<th>(df)</th>
<th>(F_{crit})</th>
<th>(F/Fc)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ADDS</td>
<td>.89</td>
<td>33</td>
<td>3</td>
<td>12</td>
<td>5.95</td>
<td>5.53</td>
</tr>
<tr>
<td>CDH</td>
<td>.81</td>
<td>24</td>
<td>2</td>
<td>13</td>
<td>6.70</td>
<td>3.55</td>
</tr>
<tr>
<td>Telecomm</td>
<td>.88</td>
<td>20</td>
<td>4</td>
<td>11</td>
<td>5.70</td>
<td>3.43</td>
</tr>
<tr>
<td>Structures</td>
<td>.76</td>
<td>20</td>
<td>2</td>
<td>13</td>
<td>6.70</td>
<td>3.00</td>
</tr>
<tr>
<td>Mech BU</td>
<td>.90</td>
<td>59</td>
<td>2</td>
<td>13</td>
<td>6.70</td>
<td>8.77</td>
</tr>
<tr>
<td>Power Gen.</td>
<td>.95</td>
<td>37</td>
<td>2</td>
<td>13</td>
<td>6.70</td>
<td>5.52</td>
</tr>
<tr>
<td>Thermal</td>
<td>.74</td>
<td>17</td>
<td>2</td>
<td>12</td>
<td>6.93</td>
<td>2.48</td>
</tr>
<tr>
<td>Propulsion</td>
<td>.93</td>
<td>90</td>
<td>2</td>
<td>13</td>
<td>6.70</td>
<td>13.48</td>
</tr>
<tr>
<td>Average</td>
<td>.85</td>
<td>29.9</td>
<td>2.4</td>
<td>12.4</td>
<td>6.5</td>
<td>4.6</td>
</tr>
<tr>
<td>Min</td>
<td>.74</td>
<td>17.2</td>
<td>2.0</td>
<td>11.0</td>
<td>5.7</td>
<td>2.5</td>
</tr>
<tr>
<td>Max</td>
<td>.9</td>
<td>90.3</td>
<td>4.0</td>
<td>13.0</td>
<td>6.9</td>
<td>13.5</td>
</tr>
</tbody>
</table>

From the summary we see that all of the coefficients of multiple determination \((R^2)\) are very high (.74 or above). The \(F\) statistics are similarly high and compare well with the \(F_{crit}\) values for each of the regression estimates. For this reason, we believe that the estimates produced by the
model are accurate predictor's of the Team X estimates for missions that fall within the range of the data base parameters. In order to visually demonstrate how the model is validated against the source data itself, we show (in figure 2) a comparison of actual propulsion subsystem costs (in the data base) with the model predicted costs. The cost estimate model for this subsystem demonstrates an extremely good fit to the data ($R^2 = .95$).

This does not mean that all work on the model is complete. Other subsystem models need further refining. A great deal of fine tuning is being conducted as our continuing sessions with the cognizant engineers bring out other causal relations and parameters that need to be validated and tested. It is the goal of the cost team to achieve results such that all of the predictive equations achieve the optimum ability to predict costs within the range of the parameters.

Fig. 2. Propulsion: Actual vs. Predicted

7. Cost Model Utilization in an Interactive Environment
The cost model CER's are currently being utilized by Team X in an interactive environment that permits spacecraft designers to see the cost impact of their design decisions as they progress. This permits them to make the necessary trades between, science, technology, and engineering practice to achieve a design that falls within a specific cost cap. Leigh Rosenberg will provide more details of this
process in his paper.

8. Concluding Remarks
The Unmanned Spacecraft Subsystem Cost Estimation Model, has evolved into one of the key tools being used to plan and cost advanced missions. The ability to predict what the Team X group of experts would estimate as the cost of a proposed mission is of great value in performing cost trades and off-line studies before calling a Team X session. Besides avoiding unnecessary planning costs, the model permits the cost analyst supporting the Team X sessions to evaluate the costs that are currently being estimated against the model. He may then bring any inconsistencies to the attention of the Team lead and have the issue resolved during the session. In every respect, the model will enhance the efficiency of the planning process and improve the quality of cost estimates for advance projects under study by Team X.

In the future, the model will also be validated against actual project implementation costs as these occur. Once a sufficient number of these new projects have been implemented and the model is modified to reflect these data, the model will become the de facto tool for predicting future project costs which are compliant to the DNP approach.

Current work on the model includes adapting the model to handle non-DNP projects and the addition of a monte carlo simulation feature.

References


Author Biography

George Madrid is a Senior Engineer in the Advanced Mission Operations Section of the Jet Propulsion Laboratory. Mr. Madrid holds a BA Degree in Mathematics from San Francisco State University, a Masters Degree in Systems Engineering from West Coast University, and Certificates in Astrodynamics and Space Communication Systems from UCLA. His technical experience has been in the
area of Orbit Determination, Navigation, and Very Long Baseline Interferometry. He has held various engineering and management positions at JPL, the latest being the manager of the Real-Time Operations Element in the Cassini Mission to Saturn. As part of his past experience, he has been involved in both software and hardware cost estimation and is a past member of the International Society for Parametric Estimation (ISPA), Project Management Institute (PMI), and the American Institute for Aeronautics and Astronautics (AIAA).